

CLAIMS

What is claimed is:

1. A machine implemented method of converting a source character string to a target character string, comprising:
 - receiving a first character string having the source character string;
 - dividing the first character string into a plurality of sub-strings;
 - converting the plurality of sub-strings to second character strings, through a dictionary;
 - creating third character strings corresponding to the plurality of sub-strings;
 - analyzing the second character strings and the third character strings;
 - constructing fourth character strings from the second and third character strings based on the analysis;
 - creating a candidate list based on the fourth character strings;
 - selecting the target character string from the candidate list; and
 - outputting the target character string;
2. The method of claim 1, wherein the first character string comprises Japanese hiragana characters and wherein the dictionary is stored in a machine readable form prior to the receiving.

3. The method of claim 1, wherein the second character strings comprise Japanese words.
4. The method of claim 1, wherein the third character strings comprise katakana characters.
5. The method of claim 1, further comprising:
constructing the fourth character strings from the second character strings, if the second character strings contain a character string corresponding to the first character string; and
constructing the fourth character strings from the third character strings if the second character strings do not contain the character string corresponding to the first character string.
6. The method of claim 1, further comprising prioritizing the second character strings and the third character strings, based on priority information associated with the second character strings and the third character strings.
7. The method of claim 6, the priority information comprises:
usage frequency of each of the second character strings;
character segment length; and
connection relationship information between the second character strings.

8. The method of claim 6, further comprising calculating a priority value for each of the second and third character strings based on the priority information.
9. The method of claim 1, further comprising retrieving priority information of the second character strings from the dictionary.
10. The method of claim 1, wherein the dictionary comprises multiple dictionaries.
11. The method of claim 1, wherein the analysis of the second and third character strings is performed based on a set of rules.
12. The method of claim 11, wherein the set of rules comprises information of usage frequency of the character strings and connection relationship between the character strings.
13. The method of claim 12, wherein the set of rules further comprises information of character segment length and semantic/grammar rules.
14. The method of claim 11, wherein the construction of the fourth character strings from the second and third character strings is based on the set of rules.
15. The method of claim 1, further comprising calculating priority value of each of the fourth character strings.

16. The method of claim 15, wherein the candidate list is created based on the priority values of the fourth character strings.
17. The method of claim 16, wherein the candidate with the highest priority value is selected as the target character string from the candidate list.
18. The method of claim 1, further comprising:
- examining the output of the target character string to determine if the target character string is correct;
 - retrieving the candidate list of alternative character strings, if the output is incorrect; and
 - selecting a character string as final output from the candidate list.
19. The method of claim 18, wherein a user examines the output of the target character string.
20. The method of claim 18, further comprising displaying the candidate list to a user interface.
21. The method of claim 18, wherein a user selects a character string as final output from the candidate list.

22. The method of claim 18, further comprising updating a database regarding to the priority information of the final output character string, based on the selection of the final output from the candidate list.
23. The method of claim 20, wherein the user interface comprises a pop-up window.
24. The method of claim 18, wherein if the candidate list does not contain the correct target character string, the method further comprises:
providing the target character string by a user;
converting the first character string to the target character string manually;
and
storing the target character string provided by the user, to a database.
25. The method of claim 24, wherein the target character string is entered by the user through a user interface.
26. The method of claim 11, wherein the set of rules is defined by a user through a user interface.
27. The method of claim 1, wherein the creation of the third character strings is performed through a look-up table.

28. The method of claim 1, wherein the creation of the third character strings is performed through an algorithm.
29. The method of claim 8, further comprising assigning a priority value to each of the third character strings.
30. The method of claim 29, wherein the priorities of the third character strings are lower than the priorities of the second character strings.
31. A data processing system implemented method for converting a first Japanese character input string to a second Japanese character string, the method comprising:
- receiving a hiragana input;
 - automatically determining a plurality of possible katakana candidates based on the hiragana input;
 - analyzing the plurality of possible katakana candidate to convert the hiragana input to katakana characters;
 - selecting one of the katakana candidates in response to the analyzing;
 - outputting converted text comprising the one of the katakana candidates;
32. The method of claim 31, wherein the converted text further comprises kanji characters.

33. An apparatus of converting a source character string to a target character string, comprising:
- means for receiving a first character string having the source character string;
 - means for dividing the first character string into a plurality of sub-strings;
 - means for converting the plurality of sub-strings to second character strings, through a dictionary;
 - means for creating third character strings corresponding to the plurality of sub-strings;
 - means for analyzing the second character strings and the third character strings;
 - means for constructing fourth character strings from the second and third character strings based on the analysis;
 - means for creating a candidate list based on the fourth character strings;
 - means for selecting the target character string from the candidate list; and
 - means for outputting the target character string;
34. The apparatus of claim 33, wherein the first character string comprises Japanese hiragana characters and wherein the dictionary is stored in a machine readable form prior to the receiving.
35. The apparatus of claim 33, wherein the second character strings comprise Japanese words.

36. The apparatus of claim 33, wherein the third character strings comprise katakana characters.
37. The apparatus of claim 33, further comprising:
- means for constructing the fourth character strings from the second character strings, if the second character strings contain a character string corresponding to the first character string; and
 - means for constructing the fourth character strings from the third character strings if the second character strings do not contain the character string corresponding to the first character string.
38. The apparatus of claim 33, further comprising means for prioritizing the second character strings and the third character strings, based on priority information associated with the second character strings and the third character strings.
39. The apparatus of claim 38, the priority information comprises:
- usage frequency of each of the second character strings;
 - character segment length; and
 - connection relationship information between the second character strings.

40. The apparatus of claim 39, further comprising means for calculating a priority value for each of the second and third character strings based on the priority information.
41. The apparatus of claim 33, further comprising means for retrieving priority information of the second character strings from the dictionary.
42. The apparatus of claim 33, wherein the dictionary comprises multiple dictionaries.
43. The apparatus of claim 33, wherein the analysis of the second and third character strings is performed based on a set of rules.
44. The apparatus of claim 43, wherein the set of rules comprises information of usage frequency of the character strings and connection relationship between the character strings.
45. The apparatus of claim 44, wherein the set of rules further comprises information of character segment length and semantic/grammar rules.
46. The apparatus of claim 43, wherein the construction of the fourth character strings from the second and third character strings is based on the set of rules.

47. The apparatus of claim 33, further comprising means for calculating priority value of each of the fourth character strings.
48. The apparatus of claim 47, wherein the candidate list is created based on the priority values of the fourth character strings.
49. The apparatus of claim 48, wherein the candidate with the highest priority value is selected as the target character string from the candidate list.
50. The apparatus of claim 33, further comprising:
- means for examining the output of the target character string to determine if the target character string is correct;
 - means for retrieving the candidate list of alternative character strings, if the output is incorrect; and
 - means for selecting a character string as final output from the candidate list.
51. The apparatus of claim 50, wherein a user examines the output of the target character string.
52. The apparatus of claim 50, further comprising means for displaying the candidate list to a user interface.

53. The apparatus of claim 50, wherein a user selects a character string as final output from the candidate list.
54. The apparatus of claim 50, further comprising means for updating a database regarding to the priority information of the final output character string, based on the selection of the final output from the candidate list.
55. The apparatus of claim 52, wherein the user interface comprises a pop-up window.
56. The apparatus of claim 50, wherein if the candidate list does not contain the correct target character string, the apparatus further comprises:
 - means for providing the target character string by a user;
 - means for converting the first character string to the target character string manually; and
 - means for storing the target character string provided by the user, to a database.
57. The apparatus of claim 56, wherein the target character string is entered by the user through a user interface.
58. The apparatus of claim 43, wherein the set of rules is defined by a user through a user interface.

59. The apparatus of claim 33, wherein the creation of the third character strings is performed through a look-up table.
60. The apparatus of claim 33, wherein the creation of the third character strings is performed through an algorithm.
61. The apparatus of claim 40, further comprising means for assigning a priority value to each of the third character strings.
62. The apparatus of claim 61, wherein the priorities of the third character strings are lower than the priorities of the second character strings.
63. A apparatus for converting a first Japanese character input string to a second Japanese character string, the apparatus comprising:
means for receiving a hiragana input;
means for automatically determining a plurality of possible katakana candidates based on the hiragana input;
means for analyzing the plurality of possible katakana candidate to convert the hiragana input to katakana characters;
means for selecting one of the katakana candidates in response to the analyzing;
outputting converted text comprising the one of the katakana candidates;

64. The apparatus of claim 63, wherein the converted text further comprises kanji characters.
65. A machine readable medium having stored thereon executable code which causes a machine to perform a method of converting a source character string to a target character string, the method comprising:
- receiving a first character string having the source character string;
 - dividing the first character string into a plurality of sub-strings;
 - converting the plurality of sub-strings to second character strings, through a dictionary;
 - creating third character strings corresponding to the plurality of sub-strings;
 - analyzing the second character strings and the third character strings;
 - constructing fourth character strings from the second and third character strings based on the analysis;
 - creating a candidate list based on the fourth character strings;
 - selecting the target character string from the candidate list; and
 - outputting the target character string;
66. The machine readable medium of claim 65, wherein the first character string comprises Japanese hiragana characters and wherein the dictionary is stored in a machine readable form prior to the receiving.

67. The machine readable medium of claim 65, wherein the second character strings comprise Japanese words.
68. The machine readable medium of claim 65, wherein the third character strings comprise katakana characters.
69. The machine readable medium of claim 65, wherein the method further comprises:
 - constructing the fourth character strings from the second character strings, if the second character strings contain a character string corresponding to the first character string; and
 - constructing the fourth character strings from the third character strings if the second character strings do not contain the character string corresponding to the first character sting.
70. The machine readable medium of claim 65, wherein the method further comprises prioritizing the second character strings and the third character strings, based on priority information associated with the second character strings and the third character strings.
71. The machine readable medium of claim 70, the priority information comprises:
 - usage frequency of each of the second character strings;
 - character segment length; and

connection relationship information between the second character strings.

72. The machine readable medium of claim 70, wherein the method further comprises calculating a priority value for each of the second and third character strings based on the priority information.
73. The machine readable medium of claim 65, wherein the method further comprises retrieving priority information of the second character strings from the dictionary.
74. The machine readable medium of claim 65, wherein the dictionary comprises multiple dictionaries.
75. The machine readable medium of claim 65, wherein the analysis of the second and third character strings is performed based on a set of rules.
76. The machine readable medium of claim 75, wherein the set of rules comprises information of usage frequency of the character strings and connection relationship between the character strings.
77. The machine readable medium of claim 76, wherein the set of rules further comprises information of character segment length and semantic/grammar rules.

78. The machine readable medium of claim 75, wherein the construction of the fourth character strings from the second and third character strings is based on the set of rules.
79. The machine readable medium of claim 65, wherein the method further comprises calculating priority value of each of the fourth character strings.
80. The machine readable medium of claim 79, wherein the candidate list is created based on the priority values of the fourth character strings.
81. The machine readable medium of claim 80, wherein the candidate with the highest priority value is selected as the target character string from the candidate list.
82. The machine readable medium of claim 65, wherein the method further comprises:
 - examining the output of the target character string to determine if the target character string is correct;
 - retrieving the candidate list of alternative character strings, if the output is incorrect; and
 - selecting a character string as final output from the candidate list.

83. The machine readable medium of claim 82, wherein a user examines the output of the target character string.
84. The machine readable medium of claim 82, wherein the method further comprises displaying the candidate list to a user interface.
85. The machine readable medium of claim 82, wherein a user selects a character string as final output from the candidate list.
86. The machine readable medium of claim 82, wherein the method further comprises updating a database regarding to the priority information of the final output character string, based on the selection of the final output from the candidate list.
87. The machine readable medium of claim 84, wherein the user interface comprises a pop-up window.
88. The machine readable medium of claim 82, wherein if the candidate list does not contain the correct target character string, the method further comprises:
 - providing the target character string by a user;
 - converting the first character string to the target character string manually;
 - and
 - storing the target character string provided by the user, to a database.

89. The machine readable medium of claim 88, wherein the target character string is entered by the user through a user interface.
90. The machine readable medium of claim 75, wherein the set of rules is defined by a user through a user interface.
91. The machine readable medium of claim 65, wherein the creation of the third character strings is performed through a look-up table.
92. The machine readable medium of claim 65, wherein the creation of the third character strings is performed through an algorithm.
93. The machine readable medium of claim 72, wherein the method further comprises assigning a priority value to each of the third character strings.
94. The machine readable medium of claim 93, wherein the priorities of the third character strings are lower than the priorities of the second character strings.
95. A machine readable medium for converting a first Japanese character input string to a second Japanese character string, the method comprising:
 - receiving a hiragana input;

automatically determining a plurality of possible katakana candidates based
on the hiragana input;
analyzing the plurality of possible katakana candidate to convert the hiragana
input to katakana characters;
selecting one of the katakana candidates in response to the analyzing;
outputting converted text comprising the one of the katakana candidates;

96. The machine readable medium of claim 95, wherein the converted text further
comprises kanji characters.
97. An apparatus for converting a source character string to a target character
string, comprising:
- an input method for receiving the first character string having the source
character string;
 - a regular dictionary coupled to convert the first character string to second
character strings;
 - a virtual dictionary coupled to generate third character strings based the first
character string;
 - a morphological analysis engine (MAE) coupled to the input method, the
MAE performing morphological analysis on the first character string
and converting the first character string to the target character string
based on the second and third character strings;
 - an output unit coupled to the MAE.

98. The apparatus of claim 97, wherein the input method comprises an application programming interface (API) for receiving the first character string.
99. The apparatus of claim 97, wherein the input method receives the first character string from a keyboard.
100. The apparatus of claim 97, wherein the input method receives the first character string from a personal digital assistant (PDA).
101. The apparatus of claim 97, wherein the first character string comprises Japanese hiragana characters.
102. The apparatus of claim 97, wherein the target character string comprises Japanese characters, wherein the Japanese characters comprise:
hiragana characters;
katakana characters; and
kanji characters.
103. The apparatus of claim 97, wherein the MAE divides the first character string into a plurality of sub-strings and converts the plurality of sub-strings to the second character strings through the regular dictionary.

104. The apparatus of claim 97, wherein the MAE divides the first character string into a plurality of sub-strings and generates the third character strings corresponding to the plurality of sub-strings through the virtual dictionary.
105. The apparatus of claim 104, wherein the virtual dictionary provides part-of-speech information for each of the third character strings.
106. The apparatus of claim 104, wherein the virtual dictionary assigns a priority value to each of the third character strings.
107. The apparatus of claim 106, wherein the priorities of the third character strings are lower than the priorities of the character strings from the regular dictionary.
108. The apparatus of claim 97, wherein the regular dictionary comprises multiple dictionaries.
109. The apparatus of claim 104, wherein the MAE selects candidates from the output of the regular dictionary, and the MAE selects the candidates from the output of the virtual dictionary if the regular dictionary does not contain the correct conversion.

110. The apparatus of claim 97, further comprising a dictionary management module (DMM) coupled to the MAE, the dictionary management managing the regular dictionary and the virtual dictionary.
111. The apparatus of claim 110, wherein the MAE accesses the regular dictionary and virtual dictionary through the DMM.
112. The apparatus of claim 97, further comprising a user interface coupled to provide user interaction to the conversion generated from the MAE.
113. The apparatus of claim 112, wherein if the conversion generated from the MAE is incorrect, a user selects a final output through the user interface, from a candidate list generated by the MAE.
114. The apparatus of claim 97, wherein the output unit transmits the conversion to a display device.
115. The apparatus of claim 97, wherein the output unit transmits the conversion to an application through an API.
116. A computer system having an apparatus according to claim 97.

117. The computer system of claim 116, wherein the computer system is a network computer and wherein the regular dictionary is stored in a network storage location.